Utrecht University

School of Economics

Group A - Team 3

A Replication Study to Evaluate the Effects of Cash Incentives Program on Low-Performing Highschool Students in Israel

Abstract

Education is one of the fundamental factors driving students' success, which leads educators, and governments to exert substantial efforts on improving their academic performance. Our paper extends the research by Angrist and Lavy (2009) on the effectiveness of cash incentives, a student-performance boosting tool, on low-performing high school students in Israel. We provide empirical evidences on their limited effects and elucidate why mainly female students with financial constraints and previous good performance are most responsive to the cash offering program. Our findings suggest a better way to evaluate and construct measures to enhance educational attainment.

Course: Data Science Lab (ECB3DSL), Academic year 2021/2022

Team: 3

[Nguyen Thuy Duong, 6827209, d.nguyenthuyduong@students.uu.nl]

[Lan Nguyen, 6687547, m.nguyenhoangmylan@students.uu.nl]

[Lisa Verlare, 6929699, l.verlare@students.uu.nl]

Supervisor(s): Ulrich Zierahn

Word count: [4726]

Statement of Originality

We, the aforementioned students, herewith declare to have written this document and that we are responsible for the content of it. We declare that the text and the work presented in this document is original and that no sources other than those mentioned in the text and its references have been used in creating it.

Utrecht University School of Economics is responsible solely for the supervision of completion of the work, not for the content.

Division of Work

We, the aforementioned students, herewith declare that we have divided the work on this project and this project paper as stated in the following table:

Section and Content	**Name and Student Number**
1 Introduction	Nguyen Thuy Duong (6827209) Lan Nguyen (6687547) Lisa Verlare (6929699)
2 Bagrut System & RCT	Lan
2.1. Bagrut System	
2.2. RCT	
3 Data Explanatory Analysis	
3.1. Data Preparation	Lisa
3.2. Descriptive Data	Duong
4 Internal Validity	Lisa
5 Empirical Approach	All (writing & coding)
5.1. Model replication	Lisa & Lan
5.2. Heterogeneity check	
5.2.1. Interaction term	Lan
5.2.2. Splitting model	Duong
6. Results	All
7 Conclusions	All
Appendices	Lisa

Signatures

Nguyen Thuy Duong (6687547)	Lisa Verlare (6929699)	Lan Nguyen (6687547)

Table of Contents

List of Tables

List of Figures

1 Introduction

The empirical research on whether monetary incentives contribute to higher academic achievements among students has yet to reach a conclusion. Studies by Vi-Nhuan (2020) and Sharma (2010) signal the positive effects of cash interventions on students' performance, while others show that these external forces could also generate negative results (Leuven et al., 2010). There is a range of factors underlying such inconsistencies, one of which could be the financial background of the students.

Angrist and Lavy (2009) suggest a relationship between students' low achievements and their family's credit constraints, which could partly explain why offering cash worked under certain circumstances but not in others. To show this, they investigate the effects of the Achievement Awards – a program offering cash to promote college attendance rates among low-achieving students in Israel. They use the completion rate of Bagrut – a matriculation certificate obtained when Israeli high school students pass a series of different subjects – as a requirement for college entrance eligibility. The authors, however, place larger attention on how cash interventions affect boys and girls differently instead of making further inferences about financial problems. From this perspective, we aim to extend their study by additionally analyzing the link between students' achievement and financial constraints.

We use the number of siblings a student has as a proxy for their families' financial hardship, assuming the larger the family size, the greater the chance of financial constraints, which might lead to a higher need to split resources among children. Our assumption is in line with the resource dilution hypothesis coined by Blake (1981, 1985) which suggests an inverse relationship between sibling size and academic achievement. Downey (2001) and Jæger (2008) confirm this negative relationship between family size and academic performance, which suggests that more siblings indicate fewer financial and parental resources available per sibling. Additionally, Steelman and Powell (1989) show a negative relationship between sibling size and the financial aid of parents, which causes a positive relationship between sibling size and other non-parental financial resources. This could lead to students working more during high school, which could have negative effects as Eckstein and Wolpin (1999) propose a negative relationship between work and academic performance.

So far, the program has proved mainly beneficial for girls, hence, we suggest reconstructing by reworking it to cater to girls with large families in the top marginal group only. Our analysis, hence, supplements the paper by Angrist and Lavy (2009) by evidencing that family size does not necessarily make a substantial contribution to students' academic achievements. More importantly, it speaks to current literature in the sense that cash incentives might not have strong effects as expected. Given the scope of the research, our findings are principally useful for Israeli educators in considering whether to prolong or extend the scope of the program.

The paper starts by providing some additional background information on the Bagrut system in Israel. Section 3 elaborates on data preparation and data statistics description. We discuss the internal validity of the study in Section 4. In Section 5 & 6, we reproduce the authors' main findings and thoroughly examine the heterogeneous relationship between the sibling groups and their financial backgrounds. The conclusions are presented in Section 7.

2 Bagrut system & RCT

2.1 The Bagrut system

The Bagrut matriculation certificate, i.e., diploma, which is awarded to Israeli high school graduates, is a official prerequisite for higher education (Angrist & Lavy, 2009). The program includes compulsory subjects, to which English and Math are given a higher weight, and school-dependent subjects. There are three main types of schools, namely state, Jewish, and Arab schools. The differences among these schools rely on religious scriptures and language subjects. All subjects are available at different levels of difficulty. In order to achieve the Bagrut, the student must pass at least one elective at the highest difficulty level and achieve at least 21 cumulative units across all other subjects (Campus Studies, n.d.). The achievement of Bagrut opens more chances for students to achieve higher education.

2.2 School-based randomization experiment (RCT)

Identification problem

Simply regressing Bagrut completion status on treatment will not result in pure causal effects. As the model is principally formulated at the school level, identification problems, particularly omitted variables bias, might arise due to missing control variables both at individual level and school level. While the authors have already controlled for several factors that are unique to each student, they failed to account for a family's financial status. Students, from low-income families, for instance, are more likely to strive for monetary rewards. Our aforementioned arguments on the resource dilution hypothesis, where more siblings indicate lower academic performance, suggest a downward bias, implying OLS underestimates the treatment effect. Hence, the authors circumvented this by running an RCT, where the identifying assumption is that both control and treatment groups are randomly assigned, and thus on average the observed (and unobserved) variables are the same.

RCT experiment design

The Achievement Awards program implemented in 2000 concerns financial incentives to stimulate student effort. Particularly, cash rewards are directly tied to performance on Bagrut subjects' exams and on Bagrut success. This incentive program aims at low-achieving students from 40 high schools with the lowest 1999 Bagrut completion rates. Here pair-matching technique was used to form pairs of participating schools, within which treatment was randomly assigned yet only the treated group had been informed about the program (Angrist & Lavy, 2009).

3 Data Explanatory Analysis

3.1 Data Preparation

The data provided by Angrist and Lavy (2009) originates from the Israeli Ministry of Education over the period 1999 to 2002. The RCT data from 2001 is of interest.

Data is provided in separate yearly datasets. They are inspected for tidiness and missing values and passed. After which they are merged, to more efficiently prepare the data. This is possible because each dataset had identical column variables.

Now merged into one dataset, we are cleaning up the data. We rename variables and entries for the variable *year* for understandability. Additional variables, e.g., *siblings_4 and ls_#,* are coded to replicate those variables that the authors use in their models. For analysis purposes, we code a categorical variable *sib_category*, which we will use in our research. Our last action in the fully merged dataset is to drop the irrelevant variables that we will not be using. We then split up the dataset into different years and create lagged quantiles and marginal groups for each dataset for further analysis.

Variable Description

A description of variables can be read in Table 1. Further analysis of these variables is provided in Section 3.2.

The *readme* file provided by the authors indicates that variables provided are numeric. We have confirmed that this is the case. Variables provided are either dummies or integers. A few have been coded to serve as factors by us.

Table 1 shows that we have created both a categorical dummy for siblings as well as a regular dummy. In Section 5 we use *siblings_4* instead of *siblings_category*. This is due to two reasons:

1. Downey (2001) indicates no adverse effects on parental resources until the number of siblings is large, i.e., at least four. This is because the resources parents provide are quite adjustable up until then.

2. $E[siblings_4] = 3.74$ for 2001. This indicates an average (natural) threshold of at least 4 siblings and an approximately equal split. After running the models shown in Table A1 we see there might only be a difference between groups with more or less than 4 siblings (see *Appendix A*).

Therefore, we use the variable that the authors use as well: *siblings_4*.

Table 1: Variable Description

Variables	Description
treated	Treatment indicator, i.e., treated (1) and treatment (0)
pair	Treatment pair. Each treated school is paired to an untreated school that are most similar in characteristics.
siblings	Number of siblings
boy	Gender indicator, i.e., boy (1) and girl (0)
immigrant	Immigrant indicator, i.e., immigrant (1) and non-immigrant (0)
father_educ	Father's years of schooling
mother_educ	Mother's years of schooling
bagrut_status	Bagrut status, i.e., Bagrut (1) and no Bagrut (0)
lagscore	Lagged Bagrut score, based on Bagrut rates from 1999
school_religious	Indicator of Jewish religious school, i.e., Jewish religious (1) and non-Jewish religious (0)
school_arab	Indicator of Arab school, i.e., Arab (1) and non-Arab (0)
year	Year indicator
ls_#	Indicators of quantiles (25, 50, 75, 100), i.e., quantile (1) and not in the quantile (0)
siblings_4	Indicator whether the individual has 4 or more siblings, i.e. 4(+) siblings (1) and less than 4 siblings (0)
sib_group	Categorical variable for the amount of siblings, ranging from 1-4+

Note: The table describes the variables in general, they are not year specific

3.2 Descriptive Data

We first replicate the statistical description of the administrative dataset for Panel A.2001 and Panel B.2002 of Table B1 (Angrist & Lavy, 2009, p. 1391) and compare the sizes of different treatment groups in Table 2. We only perform descriptive statistics replication on the Experimental Sample, given the availability of data.

Table 2: Descriptive Statistics

	All	Boy sample	Girl sample
Panel A. 2001			
Bagrut rate	0.243	0.429	0.2
	(0.4)	(0.2)	(0.4)
School covariates			
Arab school	0.348	0.476	0.374
	(0.484)	(0.374)	(0.484)
Religious school	0.115	0.319	0.084
	(0.278)	(0.084)	(0.278)
Micro covariates			
Father's education	10.07	3.067	9.819
	(3.108)	(9.819)	(3.108)
Mother's education	10.018	3.287	9.867
	(3.321)	(9.867)	(3.321)
Number of siblings	3.742	2.662	3.651
	(2.639)	(3.651)	(2.639)
Immigrant	0.064	0.244	0.029
	(0.167)	(0.029)	(0.167)
Lagged score	53.12	29.381	52.062
	(29.421)	(52.062)	(29.421)
Siblings category	0.421	0.494	0.408
	(0.492)	(0.408)	(0.492)
Number of observations	3821	1960	1861
Panel B. 2000			
Bagrut rate	0.224	0.417	0.177
	(0.382)	(0.177)	(0.382)
School covariates			
Arab school	0.319	0.466	0.352
	(0.478)	(0.352)	(0.478)
Religious school	0.134	0.34	0.098
	(0.298)	(0.098)	(0.298)
Micro covariates			
Father's education	9.871	3.074	9.749
	(3.147)	(9.749)	(3.147)
Mother's education	9.798	3.256	9.705
	(3.326)	(9.705)	(3.326)
Number of siblings	3.682	2.466	3.531
	(2.337)	(3.531)	(2.337)
Immigrant	0.074	0.261	0.039
	(0.194)	(0.039)	(0.194)
Lagged score	50.233	28.927	49.13
	(29.368)	(49.13)	(29.368)
Siblings category	0.431	0.495	0.408
	(0.492)	(0.408)	(0.492)
Number of observations	4039	2038	2001

Note: Table columns report sample means and standard deviations (shown in parentheses). This is a replication of Table B1 (Angrist & Lavy, 2009, p. 1391)

Our principal focus is the sample data for 2001, which records the main Bagrut outcomes for the treated cohort. That is why the second half concentrates on the correlations between our variable of interest *(siblings_4)* and others in 2001.

Data Summary

In January 2001 the baseline data was collected and the outcome data for the Bagrut results originate from June 2001.

The dataset encompasses school covariates (*Arab school* and *religious school*), and a set of micro variables regarding parental education, family size, immigration status, and lagged Bagrut score. These variables capture certain socioeconomic features of the Israeli society and students in the sample at the time the research was conducted.

Sample sizes

There are 4,039 observations in the 2000 dataset, but due to the 39th school dropping out, it shrinks to 3,821 in 2001. However, the sample size remains efficiently large and there are no significant differences. Hence, the results are likely to be reliable.

Covariates

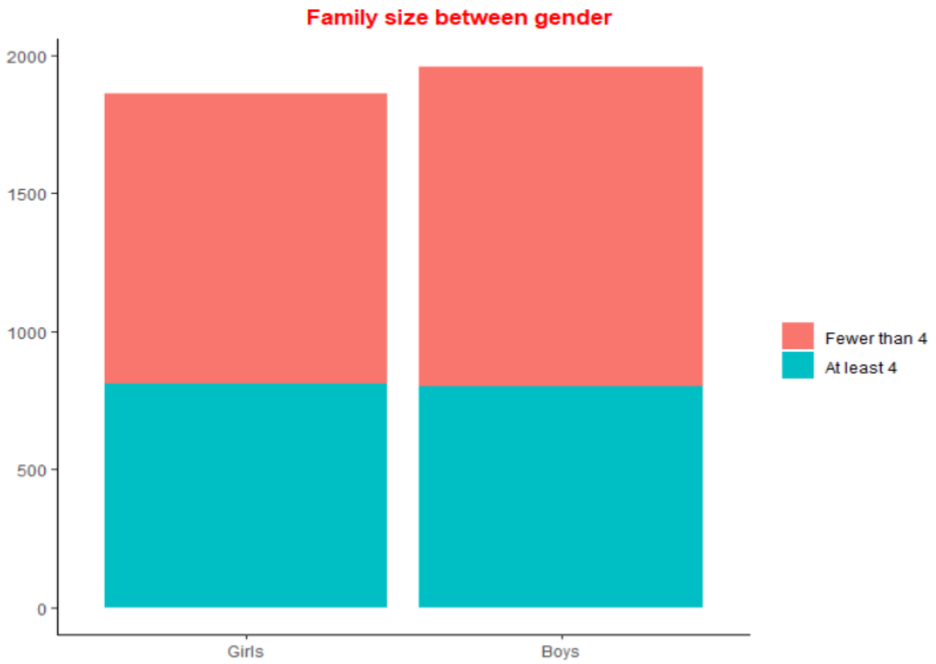
Relatively more students attend Arab schools than religious schools. On average, fathers tend to obtain 8 – 10 years of education, relatively comparable to schooling achieved by mothers. The gender distribution is almost equal, with most of them being natives (approximately 93%). All students have at least one sibling, but the frequency of having 3 siblings is the highest. Prior to the treatment, the mean of lagged Bagrut scores of girls is higher than that of boys. The difference increases after the treatment.

Multivariate analysis

Over 80% of the students attending Arab schools have a large family size, in contrast to those enrolled in religious schools (roughly 48%). Remarkably, parents in a bigger family tend to have less education, but differences are trivial when they are either relatively low educated (0 – 7 years) or highly educated (13(+) years).

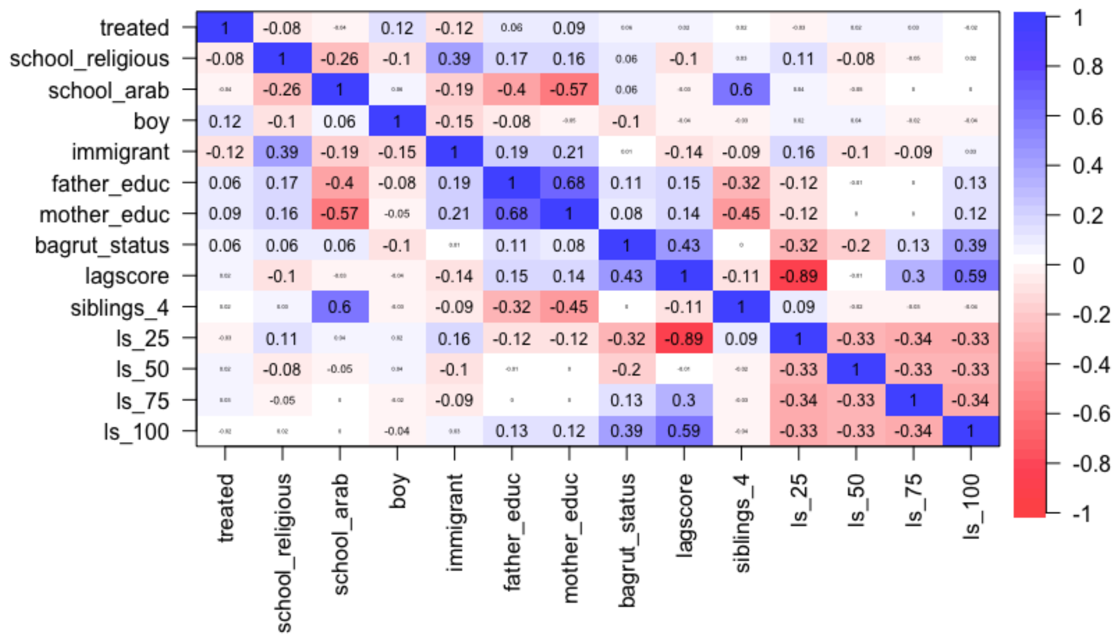
Family size varies between natives and immigrants, but most students tend to have fewer than four siblings. However, given the number of immigrant observations, this finding should be interpreted with care. Although students with at least 4 siblings tend to have lower score performance, this can be due to the size of the two sibling categories - there are more observations for those with fewer than 4 siblings. Figure 1 indicates more boys are observed than girls, respectively 1,960 and 1,861, but there is no difference in their family sizes between the two gender categories.

Figure 1: Family size between gender



The correlation matrix in Figure 2 reports statistical relationships between the number of siblings and other parameters. As shown, there is a moderately negative correlation between the number of siblings and parental education levels. The correlation between the number of siblings and lagged scores of students is weakly negative (-0.1065). Strikingly, the Bagrut status is uncorrelated to the student family size. In contrast, cash intervention is unassociated with family size.

Figure 2: Correlation Matrix 2001



4 Internal Validity

Incomplete randomization

Random assignment is very important in an RCT to get causal estimates (Kendall, 2003). Thus, complete randomization would be an indication of a good RCT. Here, in Table 3, we can see whether there are significant differences between those in the treatment and control groups. We assume that the two groups are not randomly assigned if there are significant differences and cannot assume they are on average the same. We have replicated Table B3 (Angrist & Lavy, 2009, p. 1410) for the baseline and treatment year, where the authors have checked for incomplete randomization.

The differences between the treatment and control groups are insignificant[1]. It does not appear there is self-selection of individuals into treatment, otherwise, the covariates would have had significant differences[2]. We can now safely assume that those in the control and treatment groups are the same on average, and thus treatment is randomly assigned. Consequently, this also indicates that on average there is no omitted variable bias. It does not appear there is self-selection of individuals into treatment, otherwise, the covariates would have had significant differences.

Incomplete treatment

While we might be able to estimate causal effects, the effect we are estimating depends on compliance. Angrist and Lavy (2009) indicate that while they gathered data on 39 schools, five schools were non-compliant. The authors define non-compliance as schools that were not transparent about the experiment to their students and teachers or did not let researchers know that they will not be participating. Therefore, we will not be able to estimate the LATE effect. Instead, the authors and this paper attempt to estimate the ITT, as there is incomplete treatment in the sample.

Hawthorne effect

Angrist and Lavy (2009) indicate that schools in the treatment group were notified they were in the experiment and that schools had to inform their students and teachers. This could lead to a Hawthorne effect among students and/or teachers because they now know their results are being evaluated. It could cause changes in the behavior of students, i.e., they are going to put

[1] We consider a result insignificant when the coefficient is smaller than 2 times the standard errors.
[2] For example: For 2000, variable Arab School has a difference of 0.05, while the standard error is 0.184. 0.05 < 0.368, thus insignificant.

in more effort because they know that if they put in more effort, they are more likely to receive the cash.

Table 3: Covariate Balance

	All		Boys		Girls	
	Mean	Differences	Mean	Differences	Mean	Differences
Year: 2001						
School covariates						
Arab school	0.348	-0.034 (0.191)	0.374	-0.147 (0.202)	0.320	0.071 (0.198)
Religious school	0.115	-0.052 (0.096)	0.084	0.093 (0.076)	0.148	-0.19 (0.137)
Micro covariates						
Father's education	10.070	0.365 (0.698)	9.819	1.313 (0.875)	10.335	-0.49 (0.631)
Mother's education	10.018	0.587 (0.872)	9.867	1.445 (1.029)	10.177	-0.219 (0.839)
Number of siblings	3.742	0.097 (0.733)	3.651	-0.11 (0.748)	3.839	0.362 (0.784)
Immigrant	0.064	-0.059 (0.072)	0.029	0.019 (0.015)	0.100	-0.126 (0.12)
Lagged score	53.120	1.168 (4.512)	52.062	-0.223 (4.856)	54.233	3.184 (6.318)
Year: 2000						
Bagrut rate	0.224	0.048 (0.055)	0.177	0.041 (0.053)	0.272	0.083 (0.072)
School covariates						
Arab school	0.319	-0.032 (0.181)	0.352	-0.131 (0.196)	0.286	0.05 (0.184)
Religious school	0.134	-0.029 (0.106)	0.098	0.096 (0.092)	0.170	-0.139 (0.149)
Micro covariates						
Father's education	9.871	0.328 (0.716)	9.749	1.271 (0.921)	9.995	-0.557 (0.695)
Mother's education	9.798	0.536 (0.882)	9.705	1.579 (1.056)	9.893	-0.459 (0.867)
Number of siblings	3.682	0.15 (0.628)	3.531	0.015 (0.621)	3.837	0.372 (0.676)
Immigrant	0.074	-0.053 (0.067)	0.039	0.012 (0.027)	0.109	-0.102 (0.105)
Lagged score	50.233	4.481 (4.71)	49.130	5.052 (4.727)	51.355	4.628 (7.149)

Note: Table columns report means and differences between the treatment and control group. Standard errors are clustered at the school level and reported in parantheses. This is a replication of Table B3 (Angrist & Lavy, 2009, p. 1410)

Hence, it is probable that treated students will put in the additional effort they otherwise would not have. However, we do not view this as a concern, as behavioral changes are what we want to observe. If a cash incentive would not incentivize students to put in additional effort, then the experiment would be unnecessary.

Insufficient sample size

Angrist and Lavy (2009) raise the concern that the sample size is insufficient for a good balance between the treatment and control group. It could result in estimates that are not precise. However, they argue that since school-level clusters are used with matching methods between treatment and control groups, the number of clusters is common to other randomized trials that also utilize clusters.

5 Empirical approach

We start by reproducing the authors' findings in 2001 because we are interested in the treatment effects (Table B2 (Angrist & Lavy, 2009, p. 1394)), particularly in the marginal groups (Table B4 (Angrist & Lavy, 2009, p. 1398)). We then statistically determine whether *siblings_4* is a relevant variable explaining the average behavior of the outcome variable (cash incentives) by doing a heterogeneity check.

Here two approaches are used: interaction terms and sample splitting. The former method requires a hypothesis of difference in the magnitude of the effect of cash rewards on Bagrut subject to family size only, while the latter also considers the differences in the effects of all other control variables. Unless the interaction effects are significant, we will regress separate models for two groups of students. If the coefficients for other confounders are significant and considerably diverged, results from sample splitting should be used to further interpret our findings. Another reason is that it can produce more straightforward interpretations and serves as an alternative to the interaction model (Harrer et al., 2022).

5.1 Model Replication

5.1.1 Treatment Results

Angrist and Lavy (2009) have estimated treatment effects on Bagrut achievement via four models, two of which include a dummy for pair effects between schools matched between treatment and control group as shown in Table 4. The first two models only control for school covariates, while the latter two also consider lagged score quartile dummies and micro covariates.

Table 4: Model Specification

	TREATMENT	PAIR	SCHOOL COVARIATES	QUARTILE DUMMIES	MICRO COVARIATES
Model 1	x		x		
Model 2	x	x	x		
Model 3	x		x	x	x
Model 4	x	x	x	x	x

Note. Pair effects are not included for models of gender subgroups otherwise we would lose observations from same-sex schools (Angrist & Lavy, 2009).

Treatment effects have been reported under OLS estimates and logit marginal effects by using OLS regression and logit regression respectively. We approach Logit, which estimates the likelihood of achieving Bagrut certification, via the Generalized Linear Model (GLM) as an alternative to Liang and Zeger (1986)'s Generalized Estimating Equation (GEE) initially used by the authors. Since GEE is used as an extension to GLM (Pekár & Brabec, 2017), we use the same strategy of correcting for standard errors as the authors, namely Bell and McCaffrey's (2002) Biased Reduced Linearization (BRL) estimator to improve robustness.

Later we use `margins` command from R to transform estimates into marginal effects. Marginal effects are defined as "partial derivatives of the regression equation with respect to each variable in the model for each unit in the data" (An Introduction to 'Margins,' 2021). We specifically estimate the average marginal effects, which means that for each student the marginal effect is calculated, whereafter the mean is calculated by adding up all the marginal effects and dividing by the number of students (Huntington-Klein, 2020).

The interpretations of results constructed with GLM are relatively the same as those by GEE since it is an extension of GLM. The difference lies in the fact that when analyzing dependent data, the GLM does not account for correlations over time from the same subjects (Pekár & Brabec, 2017), for example, students studying in the same school for more than one year. This explains the minor differences between the original results and our replicated ones.

5.1.2 Marginal Results

Table B4 (Angrist & Lavy, 2009, p. 1398) reports the Logit marginal effects. We categorize students into two subgroups of genders and further divide them into top and bottom subgroups according to either their lagged scores or predicted probability of matriculation quantiles. For two subgroups of boys and girls, the probability of Bagrut success is separately predicted, using school covariates, micro covariates, and lagged score quartile dummies of each group in 2001 as predictors. The upper group are those in the third and fourth quartiles, and observations in the first and second quantiles go to the lower group. The *number of students* implying the number of observations in each subgroup is used to check whether we have created the right subgroups. The output generated also includes the dependent variable mean indicating the possibility of sampled students achieving their Bagrut.

We calculate two types of (no-pair) models which both include school covariates. Another confounder in the first model is a quartile dummy while the other includes either a linear lagged score or a predicted probability corresponding to their previous classification. Although the authors have proved lagged score quartile dummies to be the most predictor so far (Angrist & Lavy, 2009), implying that the higher the past scores, the more capable the student is to pass the qualification, the use of predicted probability is to quantify individual capability in exact number on the same scale to assess the robustness of findings using lagged scores. Standard errors are adjusted using the BRL method.

5.2 Heterogeneity check

5.2.1 Interaction term

Given the dummy *siblings_4* is a proxy for a family's financial status, we add the interaction term *treated*siblings_4* to further extend the four models in Table B2 (Angrist & Lavy, 2009, p. 1394). Due to the limitations of codes for logit models, the interaction effects are only reported in OLS estimates. Only if OLS interaction effects are significant will the marginal effects be reported. Here both OLS and logit regressions are used to determine the marginal effects and at the same time to check their robustness. Our results are robust if interpretations from both regressions are (in)significant.

5.2.2 Splitting model

We again use the OLS and Logit Regressions to create restricted models and apply the BRL to adjust the standard errors. These steps are performed on the whole sample and marginal groups of genders. Our main objective is to assess if financial status, which is translated into the number of siblings a student has, is attributed to differences between genders in program response.

In the final phase, we investigate the deviations of the estimated coefficients of socioeconomic features and standard errors between the restricted models. This helps to determine whether there should be separate models for marginal groups of genders.

6 Results

6.1 Treatment & Marginal results

The replicated results provided in Table 5 & Table 6 are comparable to Table B2 (Angrist & Lavy, 2009, p. 1394) and Table B4 (Angrist & Lavy, 2009, p. 1398) provided by the authors, with a small margin of errors ranging between 0.001 and 0.005 on average.

Table 5: Treatment Effects

		All		Boys		Girls	
	Pair Effects	OLS	Logit	OLS	Logit	OLS	Logit
Year: 2001							
Dependent Variable Mean		0.243	0.243	0.2	0.2	0.287	0.287
School Covariates							
Model 1	No	0.056	0.056	-0.01	-0.011	0.105	0.103
		(0.049)	(0.049)	(0.052)	(0.053)	(0.061)	(0.059)
Model 2	Yes	0.052	0.059	—	—	—	—
		(0.047)	(0.047)				
School Covariates, Quartile Dummies, Micro Covariates							
Model 3	No	0.051	0.047	-0.024	-0.024	0.104	0.098
		(0.04)	(0.04)	(0.044)	(0.044)	(0.048)	(0.044)
Model 4	Yes	0.064	0.055	—	—	—	—
		(0.037)	(0.038)				
Number of Students		3821	3821	1960	1960	1861	1861
Number of Schools		39	39	34	34	34	34
Year: 2000							
Dependent Variable Mean		0.224	0.224	0.177	0.177	0.272	0.272
School Covariates							
Model 1	No	0.05	0.05	0.045	0.045	0.075	0.073
		(0.056)	(0.056)	(0.06)	(0.061)	(0.067)	(0.065)
Model 2	Yes	0.043	0.048	—	—	—	—
		(0.059)	(0.062)				
School Covariates, Quartile Dummies, Micro Covariates							
Model 3	No	0.03	0.018	0.009	0.006	0.067	0.05
		(0.04)	(0.04)	(0.049)	(0.05)	(0.045)	(0.041)
Model 4	Yes	0.042	0.028	—	—	—	—
		(0.043)	(0.044)				
Number of Students		4039	4039	2038	2038	2001	2001
Number of Schools		39	39	33	33	35	35

Note: This table reports both OLS and logit estimates, i.e., marginal effects. The estimates in this table were constructed using the samples from 2001 and 2000. BRL standard errors are reported in parantheses. This is a replication of Table B2 (Angrist & Lavy, 2009, p. 1394)

We observe that coefficients for the whole sample (Table 5) and the subgroup of boys (Tables 5 & 6) are insignificant. However, the coefficients for subgroup girls in model 3 (Table 5) are significant[3], which can be due to the strong effects of the cash interventions on the top marginal group, as evidenced in Table 6. For this sample, the program is likely to lead to an increase of 0.206 percentage points in the probability of Bagrut completion.

Table 6: Treatment Effects in Covariate Subgroups

	By lagged score				By predicted probability			
	Boys		Girls		Boys		Girls	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Year: 2001								
Dependent Variable Mean	0.365	0.035	0.518	0.056	0.368	0.032	0.518	0.056
Models with:								
School Covariates, Quartile dummies	-0.013	0.007	0.206	-0.017	-0.046	0.005	0.191	-0.013
	(0.083)	(0.018)	(0.071)	(0.019)	(0.074)	(0.019)	(0.069)	(0.019)
School Covariates, Linear Lagged Score or Predicted Prob.	-0.009	0.007	0.212	-0.018	-0.043	0.001	0.206	-0.015
	(0.083)	(0.018)	(0.071)	(0.018)	(0.076)	(0.017)	(0.07)	(0.021)
Number of Students	980	980	933	928	980	980	932	929

Note: This table reports logit estimates, i.e., marginal effects. The estimates in this table were constructed using the sample of 2001. BRL standard errors are reported in parantheses. This is a replication of Table B4 (Angrist & Lavy, 2009, p. 1398)

More importantly, the higher Bagrut mean of girls, in both tables, highlights a higher chance of matriculation, compared to the other (sub)groups. It has also increased compared to the pre-treatment year shown in Table 5, which indicates the positive effects of the intervention.

Regarding the robustness, both OLS estimates and logit marginal effects are similar with few differences, as shown Table 5. In Table 6 we only use one model; hence we have removed one of our control variables *(school_arab)* to check whether coefficients change dramatically, which they do not as seen in Table C1. From these tests we can conclude that both Table 5 & 6 produce robust results.

6.2 Interaction results

Table 7: Interaction Effects

	Pair Effects	All	Boys	Girls
School Covariates				
Model 1	No	0.0746	0.0298	0.0772
		(0.0747)	(0.0712)	(0.1011)
Model 2	Yes	0.1065	—	—
		(0.0625)		
School Covariates, Quartile Dummies, Micro Covariates				
Model 3	No	0.0622	-0.0126	0.1062
		(0.0584)	(0.0688)	(0.0722)
Model 4	Yes	0.0974	—	—
		(0.0554)		

Note: The table reports OLS estimates. The estimates in this table were constructed using the sample of 2001. BRL standard errors are reported in parentheses.

The OLS estimates of interaction effects for all models in Table 7 are positive, except for the model of boys with a full set of controls. Yet these are insignificant for all models, which might

[3] e.g., 0.104 > 0.096

be attributed to the absence of effect modification. In other words, there might be variations in other socioeconomic controls that should be accounted for between the two sibling groups.

6.3 Sample splitting

Not surprisingly, the OLS and Logit regressions both emphasize the imbalanced treatment effects on male-female because of differential financial backgrounds, implying that the results are robust. Logit estimates are interpreted as we are interested in the marginal effects of the intervention. We emphasize the treatment coefficients, but also show the estimates for socioeconomic features are different between the two sibling groups. Hence, the sample splitting model is appropriate in this case.

Whole sample

Table 8 demonstrates relatively similar program effects and small deviations in standard errors between the two sibling groups in the whole experimental dataset that do not distinguish genders. Models 1 and 3 produce insignificant coefficients and comparably low standard errors, indicating that the treatment effects are comparable between two groups of siblings. However, when taking pair effects into account in models 2 and 4, cash offering impact varies. This is evidenced by an increase in the Bagrut eligibility probability by 0.1297 percentage points for students with at least 4 siblings relative to 0.0231 percentage points for others. Similar standard errors, respectively 0.0537 and 0.0562, further confirm that such a difference is not due to noise, thus indicating that gains by students with at least 4 siblings are statistically significant.

Gender subgroups

When restricting the models to subgroups of genders, male-female differences in the response to financial incentives are detected. Table 9 suggests that boy students are unaffected by the program regardless of family sizes and pair effects as there are no significant estimates shown. Conversely, Table 10 shows a consistent pattern of slightly positive female responses to the cash interventions.

It, however, bears emphasizing that although the program seems to matter to girls only, its effects are mostly insignificant and vary between sibling groups. The estimated Bagrut rates are significant only when school and micro covariates are considered. In this setting, treated girls in a larger family are more responsive as they have a higher probability of estimated matriculation gains by 0.1602 percentage points (s.e. = 0.04984), while those living with fewer siblings have a modest, insignificant, increase of 0.0594 (s.e. = 0.0597), compared to the control group. These imbalanced effects are likely to be caused by monetary support since the standard errors of the two restricted models are equal in size.

Table 8: Sample Splitting - All

	OLS								Logit							
	Model 1		Model 2		Model 3		Model 4		Model 1		Model 2		Model 3		Model 4	
	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4
Dependent Variable Mean	0.243	0.242	0.243	0.242	0.243	0.242	0.243	0.242	0.243	0.242	0.243	0.242	0.243	0.242	0.243	0.242
Treated	0.1003	0.0326	0.0909	7e-04	0.0924	0.0273	0.1413	0.0196	0.099	0.0327	0.0815	0.0164	0.0879	0.0217	0.1297	0.0231
	(0.062)	(0.0611)	(0.0449)	(0.0592)	(0.0565)	(0.0483)	(0.0484)	(0.0604)	(0.0612)	(0.0611)	(0.0616)	(0.0572)	(0.0558)	(0.0496)	(0.0537)	(0.0562)
School Covariates																
Arab School	0.1349	0.1029	0.2105	0.1582	0.1257	0.1211	0.1148	0.1648	0.1429	0.0964	0.2006	0.146	0.1161	0.1103	0.066	0.1511
	(0.0572)	(0.0646)	(0.0627)	(0.0705)	(0.0561)	(0.051)	(0.0652)	(0.0517)	(0.066)	(0.0547)	(0.0943)	(0.0644)	(0.0546)	(0.0442)	(0.0719)	(0.0549)
Religious School	0.078	0.1953	0.1718	0.2527	0.0978	0.1734	0.115	0.2213	0.0859	0.1699	0.1807	0.2324	0.0994	0.1658	0.1012	0.213
	(0.1158)	(0.0697)	(0.0911)	(0.0586)	(0.0817)	(0.0595)	(0.094)	(0.0544)	(0.1308)	(0.0508)	(0.0939)	(0.0485)	(0.0696)	(0.0454)	(0.0985)	(0.0465)
Micro Covariates																
Father's Education	—	—	—	—	0.004	0.0059	0.0032	0.0041	—	—	—	—	0.004	0.0057	0.0032	0.0033
					(0.0055)	(0.0047)	(0.0049)	(0.0046)					(0.0046)	(0.0045)	(0.004)	(0.0043)
Mother's Education	—	—	—	—	3e-04	0.0096	0.0028	0.0103	—	—	—	—	-7e-04	0.009	0.001	0.0099
					(0.004)	(0.0051)	(0.0027)	(0.0044)					(0.0034)	(0.0051)	(0.0021)	(0.0044)
Immigrant	—	—	—	—	0.0335	0.0184	0.1757	0.0364	—	—	—	—	-0.0021	0.0539	-0.0096	0.0662
					(0.0971)	(0.048)	(0.1488)	(0.0515)					(0.1615)	(0.0379)	(0.2026)	(0.0504)
Lagged Score Quartile Dummies																
2nd	—	—	—	—	0.0997	0.0956	0.1042	0.0836	—	—	—	—	0.3707	0.4635	0.3459	0.4189
					(0.038)	(0.024)	(0.0475)	(0.0254)					(0.1029)	(0.0825)	(0.1103)	(0.0777)
3rd	—	—	—	—	0.3606	0.3214	0.3651	0.2952	—	—	—	—	0.5865	0.6836	0.5695	0.6176
					(0.0494)	(0.0388)	(0.0639)	(0.0352)					(0.0913)	(0.0817)	(0.1018)	(0.0741)
4th	—	—	—	—	0.557	0.4977	0.5604	0.4665	—	—	—	—	0.6949	0.7819	0.6681	0.7166
					(0.0609)	(0.0576)	(0.0723)	(0.0494)					(0.0911)	(0.083)	(0.1028)	(0.0746)
Number of Students	1608	2213	1608	2213	1608	2213	1608	2213	1608	2213	1608	2213	1608	2213	1608	2213

Note: This table reports both OLS and logit estimates, i.e., marginal effects. The estimates in this table were constructed using the samples from 2001. BRL standard errors are reported in parantheses.

Table 9: Sample Splitting - Boys

	OLS								Logit							
	Model 1		Model 2		Model 3		Model 4		Model 1		Model 2		Model 3		Model 4	
	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4
Dependent Variable Mean	0.181	0.213	0.181	0.213	0.181	0.213	0.181	0.213	0.181	0.213	0.181	0.213	0.181	0.213	0.181	0.213
Treated	0.0122	-0.0212	—	—	-0.0154	-0.0275	—	—	0.0126	-0.0214	—	—	-0.0125	-0.0313	—	—
	(0.0592)	(0.0605)			(0.0771)	(0.048)			(0.0614)	(0.0605)			(0.0742)	(0.0491)		
School Covariates																
Arab School	0.0297	0.1261	—	—	0.0603	0.1801	—	—	0.032	0.1147	—	—	0.047	0.1568	—	—
	(0.0565)	(0.0833)			(0.0817)	(0.0629)			(0.0629)	(0.0669)			(0.0765)	(0.0538)		
Religious School	0.1653	0.359	—	—	0.2071	0.3365	—	—	0.1401	0.2709	—	—	0.1705	0.2642	—	—
	(0.2469)	(0.0626)			(0.159)	(0.0548)			(0.1782)	(0.036)			(0.0893)	(0.0377)		
Micro Covariates																
Father's Education	—	—	—	—	-0.001	0.0016	—	—	—	—	—	—	-0.0018	5e-04	—	—
					(0.0069)	(0.0054)							(0.0056)	(0.0048)		
Mother's Education	—	—	—	—	0.0089	0.0249	—	—	—	—	—	—	0.0065	0.0231	—	—
					(0.0077)	(0.0073)							(0.0064)	(0.0068)		
Immigrant	—	—	—	—	0.1885	0.054	—	—	—	—	—	—	0.2065	0.0585	—	—
					(0.416)	(0.078)							(0.1735)	(0.0651)		
Lagged Score Quartile Dummies																
2nd	—	—	—	—	0.0993	0.058	—	—	—	—	—	—	0.3201	0.359	—	—
					(0.0385)	(0.0194)							(0.1245)	(0.1371)		
3rd	—	—	—	—	0.251	0.2713	—	—	—	—	—	—	0.4446	0.5706	—	—
					(0.0629)	(0.039)							(0.119)	(0.132)		
4th	—	—	—	—	0.4626	0.3857	—	—	—	—	—	—	0.5576	0.6365	—	—
					(0.0687)	(0.0519)							(0.1224)	(0.1308)		
Number of Students	800	1160	800	1160	800	1160	800	1160	800	1160	800	1160	800	1160	800	1160

Note: This table reports both OLS and logit estimates, i.e., marginal effects. The estimates in this table were constructed using the samples from 2001. BRL standard errors are reported in parantheses.

Table 10: Sample Splitting - Girls

	OLS								Logit							
	Model 1		Model 2		Model 3		Model 4		Model 1		Model 2		Model 3		Model 4	
	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4
Dependent Variable Mean	0.304	0.274	0.304	0.274	0.304	0.274	0.304	0.274	0.304	0.274	0.304	0.274	0.304	0.274	0.304	0.274
Treated	0.157	0.0854	—	—	0.1772	0.0626	—	—	0.1484	0.085	—	—	0.1602	0.0594	—	—
	(0.0839)	(0.0801)			(0.0625)	(0.062)			(0.0777)	(0.0785)			(0.0498)	(0.0597)		
School Covariates																
Arab School	0.2372	0.1008	—	—	0.1953	0.1191	—	—	0.2403	0.0933	—	—	0.1767	0.1096	—	—
	(0.0842)	(0.0885)			(0.0621)	(0.0496)			(0.0899)	(0.0768)			(0.0507)	(0.0399)		
Religious School	0.0731	0.0842	—	—	0.0793	0.0292	—	—	0.0613	0.0837	—	—	0.0855	0.0516	—	—
	(0.122)	(0.0657)			(0.082)	(0.0766)			(0.1722)	(0.063)			(0.0667)	(0.0796)		
Micro Covariates																
Father's Education	—	—	—	—	0.0072	0.0106	—	—	—	—	—	—	0.0082	0.0106	—	—
					(0.0061)	(0.0052)							(0.005)	(0.0053)		
Mother's Education	—	—	—	—	-0.0024	-0.0011	—	—	—	—	—	—	-0.0032	-0.0014	—	—
					(0.0043)	(0.0058)							(0.004)	(0.0058)		
Immigrant	—	—	—	—	0.1409	0.098	—	—	—	—	—	—	-1.6176	0.107	—	—
					(0.1007)	(0.0544)							(0.0961)	(0.0561)		
Lagged Score Quartile Dummies																
2nd	—	—	—	—	0.1057	0.1135	—	—	—	—	—	—	0.3726	0.4911	—	—
					(0.0365)	(0.0305)							(0.1321)	(0.1302)		
3rd	—	—	—	—	0.4561	0.3465	—	—	—	—	—	—	0.6584	0.7138	—	—
					(0.0772)	(0.0484)							(0.133)	(0.1264)		
4th	—	—	—	—	0.6368	0.5864	—	—	—	—	—	—	0.7662	0.8467	—	—
					(0.0551)	(0.0688)							(0.1251)	(0.1283)		
Number of Students	808	1053	808	1053	808	1053	808	1053	808	1053	808	1053	808	1053	808	1053

Note: This table reports both OLS and logit estimates, i.e., marginal effects. The estimates in this table were constructed using the samples from 2001. BRL standard errors are reported in parantheses.

We deepen our findings by assessing the responses of females from two family sizes at the top and bottom quantiles. Among 933 female students in the upper lagged score quantiles, the program effect is largest for those living with at least 4 other children for both models 1 and 3. Table 11 shows that under model 1 more than 56% of the girls in the top-marginal group received Bagrut status. The cash offers can increase the matriculation likelihood of this group by as high as 0.308 percentage points (s.e. = 0.07). This result does not change much as we use predicted probability quantiles, 0.275 (s.e. = 0.074). Girls in the lower marginal group are unresponsive to the treatment.

The splitting models reflect that gender differentials lead to significant differences in males'-females' reactions to financial incentives in education. The number of siblings also contributes to these differences, but its influence is too modest to explain the overall program treatment impact. These findings are consistent with those produced by the interaction term model. Nevertheless, the splitting models generate more meaningful insights by revealing that girls with financial constraints who have previously performed well benefit most.

Table 11: Sample Splitting - Marginal Girls Subgroup

	Quartile Dummies (ls)				Quartile Dummies (p)			
	Top		Bottom		Top		Bottom	
	>= 4	< 4	>= 4	< 4	>= 4	< 4	>= 4	< 4
Dependent Variable Mean	0.562	0.485	0.058	0.054	0.559	0.488	0.058	0.054
Models:								
Model 1	0.285	0.154	0.006	-0.041	0.277	0.157	0.007	-0.032
	(0.071)	(0.109)	(0.036)	(0.022)	(0.074)	(0.107)	(0.036)	(0.018)
Model 3	0.308	0.149	0.001	-0.036	0.275	0.142	0.002	-0.029
	(0.07)	(0.108)	(0.036)	(0.018)	(0.074)	(0.107)	(0.036)	(0.016)
Number of Students	395	538	413	515	397	535	411	518

Note: This table reports logit estimates, i.e., marginal effects. The estimates in this table were constructed using the samples from 2001. BRL standard errors are reported in parantheses.

7 Conclusion

In this paper, we have attempted to partly reproduce the work of Angrist and Lavy (2009), which investigates whether offering financial benefits to low-performing high school students in Israel could increase their college attendance. We focus on replicating Tables B2 and B4 where effects are most observable. Our findings confirm the authors' conclusion that the financial schemes did not affect male students, evidenced by no significant change in their Bagrut performance regardless of receiving cash or not. We also observe that only girls in the marginal top group, i.e., those that were already in the higher quartile of lagged scores, are slightly more responsive to the cash offers.

The differentials in male-female responsiveness to the treatment inspired us to extend the authors' work as we suspect that the treatment results could be influenced by the participants' financial situation. We use the number of siblings a student has as an indicator of his or her financial constraints, as suggested by previous studies (Blake, 1981; Blake, 1985; Downey, 2001; Eckstein and Wolpin, 1999, Jæger, 2008; Steelman and Powell, 1989). At first, this relationship is checked with an interaction term model, and then with sample splitting. While the first method indicates no significant results, the second detects certainly positive effects on girls. Nevertheless, the splitting model highlights that this is only strongly applicable to those living with at least 4 siblings. We also investigate the marginal group of females of top and bottom quantiles of academic performance and see that in accordance with Angrist and Lavy (2009) only those close to achieving their Bagrut will also benefit most from the cash incentives.

Our research is strongly based on the Israel Achievement Awards demonstration which was particularly designed for low-achieving groups of high school students performing a high-stakes exam. Therefore, what we concluded is only limited to the scope of this research applicable to this sample. This follows that our findings should not be considered representative outcomes for high-achieving individuals and schools, or for other countries and societies, which encompass diverged socioeconomic features.

This study also channels space for future studies. As indicated, despite the potential economic gains created by the incentive scheme, gender differential in treatment response exists. Hence, Israeli educators could enhance their program by retargeting the treatment group, for instance, focusing on girls that score in the highest quartiles from large families, who tend to face more financial burdens. Besides, since the data demonstrates skewed attendance towards Arab schools, researchers should verify if the program would generate different outcomes for students in religious schools or state schools. They could simultaneously establish a new program that can incentivize male students.

From a broader perspective, other scientists can run a longitudinal study spanning multiple countries to see if monetary interventions could generate social and economic gains and under which socioeconomic backgrounds students benefit more.

REFERENCES

An Introduction to 'margins.' (2021, January 21). The Comprehensive R Archive Network. Retrieved March 22, 2022, from https://cran.r-project.org/web/packages/margins/vignettes/Introduction.html

Angrist, J., & Lavy, V. (2009). The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial. *American Economic Review*, *99*(4), 1384–1414. https://doi.org/10.1257/aer.99.4.1384

Bell, R. M., & McCaffrey, D. F. (2002). Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples. *Survey Methodology*, *28*(2), 169–181. https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2002002/article/9058-eng.pdf?st=WVgvP5RB

Blake, J. (1981). Family Size and the Quality of Children. *Demography*, *18*(4), 421–442. https://doi.org/10.2307/2060941

Blake, J. (1985). Number of Siblings and Educational Mobility. *American Sociological Review*, *50*(1), 84–94. https://doi.org/10.2307/2095342

Campus Studies. (n.d.). *About Israeli Bagrut*. Retrieved April 6, 2022, from https://campus-studies.co.il/bagrut/about-israeli-bagrut/

Downey, D. B. (2001). Number of siblings and intellectual development: The resource dilution explanation. *American Psychologist*, *56*(6–7), 497–504. https://doi.org/10.1037/0003-066x.56.6-7.497

Eckstein, Z., & Wolpin, K. I. (1999). Why Youths Drop out of High School: The Impact of Preferences, Opportunities, and Abilities. *Econometrica*, *67*(6), 1295–1339. https://doi.org/10.1111/1468-0262.00081

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (n.d.). *Chapter 7 Subgroup Analyses | Doing Meta-Analysis in R*. Doing Meta-Analysis in R: A Hands-on Guide. Retrieved March 28, 2022, from https://bookdown.org:443/MathiasHarrer/Doing_Meta_Analysis_in_R/subgroup.html#subgroup

Huntington-Klein, N. [Nick Huntington-K. (2020, September 4). *Econometrics - Marginal Effects for Probit and Logit (and Marginal Effects in R)* [Video]. YouTube.

https://www.youtube.com/watch?v=lJMO5kqWwIo&ab_channel=NickHuntington-Klein

Jæger, M. M. (2008). Do Large Sibships Really Lead to Lower Educational Attainment? *Acta Sociologica*, *51*(3), 217–235. https://doi.org/10.1177/0001699308094167

Kendall, J. M. (2003). Designing a research project: randomised controlled trials and their principles. *Emergency Medicine Journal*, *20*(2), 164–168. https://doi.org/10.1136/emj.20.2.164

Leuven, E., Oosterbeek, H., & van der Klaauw, B. (2010). The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment. *Journal of the European Economic Association*, *8*(6), 1243–1265. https://doi.org/10.1162/jeea_a_00024

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Pekár, S., & Brabec, M. (2017). Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. *Ethology*, *124*(2), 86–93. https://doi.org/10.1111/eth.12713

Sharma, D. (2010). The Impact of Financial Incentives on Academic Achievement and Household Behavior: Evidence from a Randomized Trial in Nepal. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1681186

Steelman, L. C., & Powell, B. (1989). Acquiring Capital for College: The Constraints of Family Configuration. *American Sociological Review*, *54*(5), 844–855. https://doi.org/10.2307/2117758

Vi-Nhuan, L. (2020). Do Student-Level Incentives Increase Student Achievement? A Review of the Effect of Monetary Incentives on Test Performance. *Teachers College Record: The Voice of Scholarship in Education*, *122*(3), 1–34. https://doi.org/10.1177/016146812012200304

APPENDIX A. SIBLINGS

We performed OLS and logit analysis on the different models from Table B2 (Angrist & Lavy, 2009, p. 1394) with the additional dummy sib_category. We want to ascertain whether any of the sibling categories have a significant effect on Bagrut status.

The categories that comprise this variable are as follows:

- 1 sibling
- 2 siblings
- 3 siblings
- 4 siblings
- More than 4 siblings

Model 1: School covariates – No Pair Effect: Significance at the 5% level is only present for sib_categorys4+. It looks as though having more than four siblings reduces the Bagrut status by 0.108 compared to having only one sibling, all else equal.

Model 2: School covariates – Pair Effect: Significance at the 5% level is only present for sib_category4+. Having more than four siblings now reduces the Bagrut status by 0.134 compared to having only one sibling, all else equal.

Model 3: School covariates, quartile dummies and micro covariates – No Pair Effect: There are no significant results among the sibling categories.

Model 4: School covariates and micro covariates – Pair Effect: There are no significant results among the sibling categories.

While model 1 and 2 show some significance results among the 4+ sibling category, models 3 and 4 do not. Therefore, we choose not to use this categorical dummy, as only the 4+ category shows significant results. This allows us to distinguish between two new categories: less than four and at least four siblings.

Table A1: Treatment Effects For Sibling Categories

Table A1: Treatment Effects For Sibling Categories

		All		Boys		Girls	
		OLS	Logit	OLS	Logit	OLS	Logit
Year: 2001							
School covariates							
Model 1							
	Treated	0.06	0.059	-0.007	-0.008	0.114	0.11
		(0.049)	(0.048)	(0.05)	(0.05)	(0.061)	(0.059)
	2 Siblings	0.008	0.008	-0.003	-0.005	0	-0.001
		(0.027)	(0.031)	(0.031)	(0.035)	(0.051)	(0.057)
	3 Siblings	-0.024	-0.028	-0.009	-0.013	-0.049	-0.056
		(0.051)	(0.056)	(0.049)	(0.053)	(0.077)	(0.085)
	4 Siblings	-0.055	-0.058	-0.035	-0.039	-0.094	-0.099
		(0.052)	(0.055)	(0.046)	(0.05)	(0.085)	(0.092)
	4+ Siblings	-0.108	-0.101	-0.122	-0.112	-0.131	-0.129
		(0.051)	(0.053)	(0.052)	(0.051)	(0.083)	(0.089)
Model 2							
	Treated	0.061	0.065	—	—	—	—
		(0.044)	(0.044)				
	2 Siblings	-0.013	-0.022	—	—	—	—
		(0.027)	(0.031)				
	3 Siblings	-0.059	-0.063	—	—	—	—
		(0.046)	(0.051)				
	4 Siblings	-0.085	-0.086	—	—	—	—
		(0.049)	(0.053)				
	4+ Siblings	-0.134	-0.126	—	—	—	—
		(0.051)	(0.053)				
School covariates, quartile dummies, micro covariates							
Model 3							
	Treated	0.052	0.048	-0.024	-0.024	0.108	0.1
		(0.04)	(0.04)	(0.044)	(0.043)	(0.047)	(0.043)
	2 Siblings	-0.004	-0.003	-0.018	-0.025	0.002	0.007
		(0.02)	(0.02)	(0.026)	(0.025)	(0.034)	(0.035)
	3 Siblings	-0.007	-0.001	-0.004	-0.007	0.003	0.014
		(0.042)	(0.044)	(0.043)	(0.041)	(0.051)	(0.055)
	4 Siblings	-0.019	-0.014	-0.005	-0.009	-0.027	-0.022
		(0.036)	(0.037)	(0.044)	(0.045)	(0.041)	(0.044)
	4+ Siblings	-0.037	-0.024	-0.027	-0.026	-0.056	-0.042
		(0.037)	(0.038)	(0.04)	(0.041)	(0.044)	(0.047)
Model 4							
	Treated	0.068	0.057	—	—	—	—
		(0.036)	(0.038)				
	2 Siblings	-0.011	-0.008	—	—	—	—
		(0.017)	(0.016)				
	3 Siblings	-0.025	-0.012	—	—	—	—
		(0.036)	(0.035)				
	4 Siblings	-0.033	-0.019	—	—	—	—
		(0.034)	(0.032)				
	4+ Siblings	-0.06	-0.04	—	—	—	—
		(0.034)	(0.034)				

APPENDIX B. TABLES FROM THE ORIGINAL PAPER

In this section tables from the original paper by Angrist and Lavy (2009) are displayed.

Table B1: Original Table 1

TABLE 1—DESCRIPTIVE STATISTICS

	Experimental sample			National		
	All (1)	Boys (2)	Girls (3)	All (4)	Boys (5)	Girls (6)
Panel A. 2001						
Bagrut rate	0.243	0.200	0.287	0.629	0.574	0.678
School covariates						
Arab school	0.348	0.374	0.320	0.163	0.159	0.167
Religious school	0.115	0.084	0.148	0.170	0.154	0.184
Micro covariates						
Father's education	10.1	9.82	10.3	12.2	12.2	12.1
	(3.07)	(3.11)	(3.00)	(3.48)	(3.48)	(3.48)
Mother's education	10.0	9.87	10.2	12.0	12.0	11.9
	(3.29)	(3.32)	(3.24)	(3.42)	(3.42)	(3.42)
Number of siblings	3.74	3.65	3.84	2.97	2.91	3.03
	(2.66)	(2.64)	(2.68)	(1.95)	(1.91)	(1.98)
Immigrant	0.064	0.029	0.100	0.023	0.021	0.025
Lagged score	53.1	52.1	54.2	—	—	—
	(29.4)	(29.4)	(29.3)			
Proportion missing						
Father's education	0.144	0.168	0.118	0.124	0.128	0.120
Mother's education	0.153	0.173	0.132	0.136	0.142	0.130
Number of siblings	0.116	0.111	0.122	0.107	0.110	0.105
Observations	3,821	1,960	1,861	76,990	36,423	40,567
Panel B. 2000						
Bagrut rate	0.224	0.177	0.272	0.611	0.560	0.657
School covariates						
Arab school	0.319	0.352	0.286	0.161	0.160	0.163
Religious school	0.134	0.098	0.170	0.171	0.154	0.186
Micro covariates						
Father's education	9.87	9.75	10.0	12.1	12.1	12.0
	(3.07)	(3.15)	(2.99)	(3.56)	(3.57)	(3.56)
Mother's education	9.80	9.71	9.9	11.9	11.9	11.9
	(3.26)	(3.33)	(3.18)	(3.48)	(3.50)	(3.45)
Number of siblings	3.68	3.53	3.84	2.99	2.92	3.06
	(2.47)	(2.34)	(2.58)	(1.98)	(1.92)	(2.03)
Immigrant	0.074	0.039	0.109	0.032	0.029	0.035
Lagged score	50.2	49.1	51.4	—	—	—
	(28.9)	(29.4)	(28.4)			
Proportion missing						
Father's education	0.109	0.121	0.096	0.087	0.094	0.080
Mother's education	0.115	0.129	0.100	0.085	0.094	0.077
Number of siblings	0.101	0.105	0.098	0.103	0.107	0.100
Observations	4,039	2,038	2,001	77,241	36,484	40,757

Notes: Columns 1–3 report sample means. Standard deviations are shown in parentheses. Statistics in columns 4–6 are from the authors' tabulation of administrative data for schools with a positive Bagrut rate in 1999.

Note. Adapted from "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial", by J. Angrist & V. Lavy, 2009, American Economic Review, 99(4), p. 1391 (https://doi.org/10.1257/aer.99.4.1384). AEA.

Table B2: Original Table 2

TABLE 2—TREATMENT EFFECTS AND SPECIFICATION CHECKS

	Pair effects	Boys + girls OLS (1)	Boys + girls Logit (2)	Boys OLS (3)	Boys Logit (4)	Girls OLS (5)	Girls Logit (6)
Panel A. 2001							
Dependent variable mean		0.243		0.200		0.287	
Model with:							
School covariates	No	0.056	0.051	−0.010	−0.011	0.105	0.093
		(0.049)	(0.045)	(0.052)	(0.055)	(0.061)	(0.053)
	Yes	0.052	0.054	—	—	—	—
		(0.047)	(0.043)				
School covariates, quartile dummies, micro covariates	No	0.052	0.047	−0.022	−0.023	0.105	0.097
		(0.039)	(0.039)	(0.043)	(0.045)	(0.047)	(0.046)
	Yes	0.067	0.055	—	—	—	—
		(0.036)	(0.036)				
Number of students		3,821		1,960		1,861	
Number of schools		39		34		34	
Panel B. 2000							
Dependent variable mean		0.224		0.177		0.272	
Model with:							
School covariates	No	0.050	0.046	0.045	0.040	0.075	0.069
		(0.056)	(0.051)	(0.060)	(0.055)	(0.067)	(0.061)
	Yes	0.043	0.045	—	—	—	—
		(0.059)	(0.058)				
School covariates, quartile dummies, micro covariates	No	0.030	0.018	0.009	0.006	0.066	0.051
		(0.041)	(0.042)	(0.050)	(0.052)	(0.046)	(0.046)
	Yes	0.043	0.030	—	—	—	—
		(0.044)	(0.046)				
Number of students		4,039		2,038		2,001	
Number of schools		39		33		35	
Panel C. 2002							
Dependent variable mean		0.305		0.257		0.357	
Model with:							
School covariates	No	−0.019	−0.019	−0.026	−0.028	−0.010	−0.010
		(0.071)	(0.071)	(0.073)	(0.075)	(0.077)	(0.078)
	Yes	−0.018	−0.018	—	—	—	—
		(0.050)	(0.059)				
School covariates, quartile dummies, micro covariates	No	−0.023	−0.021	−0.026	−0.024	−0.015	−0.014
		(0.044)	(0.045)	(0.046)	(0.047)	(0.046)	(0.046)
	Yes	−0.027	−0.033	—	—	—	—
		(0.033)	(0.034)				
Number of students		4,328		2,269		2,059	
Number of schools		38		33		33	

Notes: The table reports OLS estimates and logit marginal effects. Panel A shows treatment effects. Results from 2000 and 2002 are specification checks. BRL standard errors are reported in parentheses. Pair effects are omitted from models estimated separately for boys and girls so as not to lose pairs that include single-sex (religious) schools.

Note. Adapted from "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial", by J. Angrist & V. Lavy, 2009, American Economic Review, 99(4), p. 1394 (https://doi.org/10.1257/aer.99.4.1384). AEA.

Table B3: Original Table A2

TABLE A2—COVARIATE BALANCE

	All		Boys		Girls	
	Mean (1)	Difference (2)	Mean (3)	Difference (4)	Mean (5)	Difference (6)
Panel A. 2001						
School covariates						
Arab school	0.348	−0.034 [0.191]	0.374	−0.147 [0.202]	0.320	0.071 [0.198]
Religious school	0.115	−0.052 [0.096]	0.084	0.093 [0.076]	0.148	−0.190 [0.138]
Micro covariates						
Father's education	10.1	0.365 [0.698]	9.82	1.31 [0.875]	10.3	−0.490 [0.631]
Mother's education	10.0	0.587 [0.872]	9.87	1.45 [1.03]	10.2	−0.219 [0.839]
Number of siblings	3.74	0.097 [0.733]	3.65	−0.110 [0.748]	3.84	0.362 [0.784]
Immigrant	0.064	−0.059 [0.072]	0.029	0.019 [0.015]	0.100	−0.126 [0.120]
Lagged score	53.1	1.17 [4.51]	52.1	−0.223 [4.86]	54.2	3.18 [6.32]
Observations	3,821		1,960		1,861	
Panel B. 2000						
Bagrut rate	0.224	0.048 [0.055]	0.177	0.041 [0.053]	0.272	0.083 [0.072]
School covariates						
Arab school	0.319	−0.032 [0.181]	0.352	−0.131 [0.196]	0.286	0.050 [0.184]
Religious school	0.134	−0.029 [0.106]	0.098	0.096 [0.092]	0.170	−0.139 [0.149]
Micro covariates						
Father's education	9.9	0.328 [0.716]	9.75	1.27 [0.922]	10.0	−0.557 [0.695]
Mother's education	9.8	0.536 [0.882]	9.71	1.58 [1.06]	9.9	−0.459 [0.867]
Number of siblings	3.68	0.150 [0.629]	3.53	0.015 [0.621]	3.84	0.372 [0.676]
Immigrant	0.074	−0.053 [0.067]	0.039	0.012 [0.027]	0.109	−0.102 [0.105]
Lagged score	50.2	4.48 [4.71]	49.1	5.052 [4.73]	51.4	4.63 [7.15]
Observations	4,039		2,038		2,001	

Notes: This table reports means and treatment-control differences by gender in 2001 (the treatment year) and 2000 (the pre-treatment year). Standard errors, clustered by school, are reported in brackets.

Note. Adapted from "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial", by J. Angrist & V. Lavy, 2009, American Economic Review, 99(4), p. 1410 (https://doi.org/10.1257/aer.99.4.1384). AEA.

Table B4: Original Table 4

Table 4—Estimates in Covariate Subgroups

	By lagged score				By predicted probability			
	Boys		Girls		Boys		Girls	
	Top (1)	Bottom (2)	Top (3)	Bottom (4)	Top (5)	Bottom (6)	Top (7)	Bottom (8)
Panel A. 2001								
Dependent variable mean	0.365	0.035	0.518	0.056	0.368	0.032	0.518	0.056
Models with:								
School covariates, quartile dummies	−0.013	0.007	0.206	−0.020	−0.047	0.005	0.194	−0.015
	(0.083)	(0.016)	(0.079)	(0.024)	(0.077)	(0.016)	(0.077)	(0.023)
School covariates, linear lagged score or predicted prob.	−0.009	0.007	0.213	−0.021	−0.044	0.001	0.207	−0.019
	(0.083)	(0.017)	(0.079)	(0.022)	(0.079)	(0.017)	(0.078)	(0.026)
Number of students	980	980	933	928	980	980	932	929
Panel B. 2000								
Dependent variable mean	0.318	0.035	0.475	0.068	0.320	0.033	0.478	0.066
Models with:								
School covariates, quartile dummies	0.055	−0.014	0.098	0.009	0.033	0.004	0.086	0.009
	(0.079)	(0.035)	(0.074)	(0.027)	(0.078)	(0.027)	(0.071)	(0.023)
School covariates, linear lagged score or predicted prob.	0.055	−0.014	0.094	0.007	0.010	0.000	0.089	0.007
	(0.079)	(0.035)	(0.072)	(0.026)	(0.077)	(0.028)	(0.070)	(0.024)
Number of students	1,022	1,016	1,004	997	1,021	1,017	1,002	999
Panel C. 2002								
Dependent variable mean	0.475	0.040	0.611	0.101	0.472	0.042	0.608	0.106
Models with:								
School covariates, quartile dummies	−0.018	−0.004	−0.017	−0.030	−0.029	−0.007	−0.006	−0.021
	(0.101)	(0.016)	(0.088)	(0.032)	(0.098)	(0.017)	(0.078)	(0.029)
School covariates, linear lagged score or predicted prob.	−0.008	−0.003	−0.013	−0.037	−0.032	−0.015	−0.001	−0.020
	(0.097)	(0.016)	(0.088)	(0.031)	(0.088)	(0.021)	(0.073)	(0.028)
Number of students	1,135	1,134	1,035	1,024	1,135	1,134	1,030	1,029

Notes: The table reports logit marginal effects in top and bottom subgroups, classified by lagged test scores or predicted probability of Bagrut success (as a function of lagged scores and covariates). Panel A shows treatment effects. Results from 2000 and 2002 are specification checks. BRL standard errors are reported in parentheses.

Note. Adapted from "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial", by J. Angrist & V. Lavy, 2009, American Economic Review, 99(4), p. 1398 (https://doi.org/10.1257/aer.99.4.1384). AEA.

APPENDIX C. ROBUSTNESS CHECKS

In this section robustness checks will be shown when applicable.

Table C1: Robustness Check of Table 6

	Original Model								Robustness Check							
	By lagged score				By predicted probability				By lagged score				By predicted probability			
	Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Year: 2001																
Dependent Variable Mean	0.365	0.035	0.518	0.056	0.368	0.032	0.518	0.056	0.365	0.035	0.518	0.056	0.368	0.032	0.518	0.056
Models with:																
School Covariates, Quartile dummies	-0.013	0.007	0.206	-0.017	-0.046	0.005	0.191	-0.013	-0.017	0.004	0.204	-0.022	-0.047	0.002	0.19	-0.014
	(0.083)	(0.018)	(0.071)	(0.019)	(0.074)	(0.019)	(0.069)	(0.019)	(0.077)	(0.016)	(0.076)	(0.02)	(0.069)	(0.018)	(0.067)	(0.019)
School Covariates, Linear Lagged Score or Predicted Prob.	-0.009	0.007	0.212	-0.018	-0.043	0.001	0.206	-0.015	-0.014	0.002	0.211	-0.017	-0.041	0	0.206	-0.015
	(0.083)	(0.018)	(0.071)	(0.018)	(0.076)	(0.017)	(0.07)	(0.021)	(0.077)	(0.016)	(0.074)	(0.019)	(0.073)	(0.017)	(0.069)	(0.02)
Number of Students	980	980	933	928	980	980	932	929	980	980	933	928	980	980	932	929

Note: This table reports logit estimates, i.e., marginal effects. The estimates in this table were constructed using the sample of 2001. BRL standard errors are reported in parantheses. The original model is a replication of Table B4 (Angrist & Lavy, 2009, p. 1398), while the robustness check has left out a variable.